An investigation of children's working memory capacity for task rules

Abstract: 222 words; Text: 11324 words; 46 references; 3 tables; 8 figures

This revised manuscript was submitted to Cognitive Development on 25th March 2019

(initial submission on 18th October 2018)

The work described in this article was funded by a research grant from the Leverhulme Trust (RPG-2014-414).

Abstract

Studies investigating the development of working memory typically measure children's ability to maintain declarative information (e.g., lists of words) for a short period of time. But working memory also relies on the maintenance of procedural information such as task rules to guide behavior. In comparison to children's working memory capacity for declarative information, little remains known about how children's ability to maintain and act on procedural information in working memory develops throughout childhood. For this reason, Experiments 1 and 2 investigated whether children's working memory capacity for task rules increased with age, using the increase in reaction time with the number of stimulus-response mappings as an index of working memory capacity. Children aged between 5 and 11 completed a stimulus identification task in which the number of stimulus-response rules was varied. Overall, set-size effects decreased with age, suggesting that younger children have a reduced working memory capacity for task rules. A proportional analysis of the reaction time data confirmed that age-related differences in overall reaction time cannot explain this finding. Finally, Experiment 3 demonstrated that age differences in basic capacity, rather than strategic ability, are the cause of the observed difference in working memory capacity. Hence, these results show that developmental differences in working memory capacity affect not only a child's ability to recall information, but also their ability to act accordingly.

Key words: Working memory capacity, working memory development, procedural representations, task rules.

Introduction

Working memory can be defined as the ability to maintain and manipulate information over a short period of time. Clearly, this ability plays a crucial role in guiding our behavior in an efficient and goal-directed manner. It is therefore not surprising that research has found a strong correlation between working memory performance and measures of academic achievement and intelligence in both adults (Oberauer, Schulze, Wilhelm & Süß, 2005) and children (Bayliss, Jarrold, Baddeley, Gunn & Leigh, 2005). In the laboratory, working memory performance is often assessed using complex span tasks (Case, Kurland, & Goldberg, 1982; Daneman & Carpenter, 1980), which measure a person's ability to hold in mind a number of discrete items (e.g., words or digits), whilst performing an unrelated processing task (e.g., counting or reading sentences). Such experiments typically find that working memory capacity (i.e., the number of items recalled in the correct order at the end of each trial) is severely limited (Cowan, 2001; Miller, 1956). Moreover, research with children has shown that this capacity limit increases throughout childhood (Gathercole, 1999). Given the importance of working memory in everyday life, a growing body of research investigates how and why working memory develops throughout childhood (e.g., Cowan, AuBuchon, Gilchrist, Ricker & Saults, 2011; Gaillard, Barrouillet, Jarrold & Camos, 2011).

The vast majority of this research measures children's ability to recall discrete items of information, such as words, digits or spatial locations. In this way, most previous research has measured people's ability to maintain *declarative* representations over a period of time. However, these declarative representations do not control behavior. Instead, the flexible control of behavior is thought to rely on the representation and execution of *procedural* representations (Oberauer, 2009; Brass, Liefooghe, Braem & De Houwer, 2017), which specify which actions must be performed to achieve the desired behavioral outcome. Given this, procedural representations can be described by condition-action rules (e.g., "if the traffic light is red, then stop"). In psychology experiments, people's ability to maintain and execute procedural representations has been investigated using choice reaction time tasks, in which participants are instructed to respond to one of several stimuli using a set of stimulus-response (S-R) mappings or “task rules”. In research with adults, the increase in reaction time (RT) with the number of S-R mappings (also known as Hick’s law; Hick, 1952; Hyman, 1953) is seen as evidence of a limited working memory capacity for task rules (Barrouillet, Corbin, Dagry & Camos, 2015; Gade, Druey, Souza & Oberauer, 2014; Van ‘t Wout, 2018). Importantly, RT set-size effects are only obtained for tasks that are currently relevant or operative, and not for other potentially relevant tasks that are currently not active (Hübner, Kluwe, Luna-Rodriguez & Peters, 2004; Van ‘t Wout, Lavric & Monsell, 2015). This finding confirms that the set-size effect described by Hick’s law reflects a limited capacity working memory, and that it does not apply to long-term memory retrieval. Additionally, the observation that set-size effects are typically not found when working memory load does not increase with increasing set-size (e.g. in studies involving saccades to a varying number of target locations; Kveraga, Boucher & Hughes, 2002) further suggests that set-size effects reflect a limited working memory capacity (also see Schneider & Anderson, 2011).

Compared to the extensive literature on working memory for declarative representations, very little remains known about children’s ability to maintain and act on procedural representations. Considering that the ultimate goal of working memory must be to guide behavior, and not just to retain relevant information in the absence of any action, it is important for research to focus on the maintenance of procedural information such as task rules alongside the study of maintenance of declarative representations. Hence, the experiments reported here were designed to investigate children’s working memory for procedural representations. In particular, Experiments 1 and 2 asked whether children’s ability to maintain and implement task rules improves with age. Experiment 3 was designed to further investigate the cause of this developmental difference. The study of children’s ability to maintain and implement procedural representations is important for both theoretical and practical reasons, as described below.

The first and most important reason for studying children’s ability to maintain and implement procedural representations is that there is evidence in both the developmental and adult literature suggesting that the effective representation of declarative information (e.g., ability to recall a task rule) does not always translate into the desired behavioral outcome (e.g., ability to execute that task rule). Specifically, research with adults (Duncan, Emslie, Williams, Johnson & Freer, 1996) and children (Towse, Lewis & Knowles, 2007) has found that an intact declarative representation of the rules of a task is not always accompanied by the correct implementation of those rules – a phenomenon known as "goal neglect". Similarly, Zelazo, Frye and Rapus (1996) required 3- and 4-year-olds to sort cards showing colored shapes firstly according to one dimension, and then (in the "post switch phase") according to the other dimension. In the post switch phase of this Dimensional Change Card Sort paradigm, Zelazo et al. (1996) found that, despite having explicit knowledge of the rule, many 3-year-olds were unable to act on that information.. Together, these findings clearly show a dissociation between declarative and procedural representations in working memory. Therefore, measuring working memory exclusively in terms of its declarative content might provide incomplete results at best, and misleading results at worst.

Second, as mentioned above, there is a strong relationship between working memory performance and measures of real-world attainment (e.g., Bayliss et al., 2005). However, less is known about which specific aspects or components of working memory are responsible for this correlation. For example, in terms of the relationship between working memory performance and academic achievement, it is possible that working memory for declarative and procedural representations predict different aspects of classroom performance. In support of this, Duncan, Schramm, Thompson and Dumontheil (2012) have previously shown that adults' ability to maintain procedural information (such as task rules) in working memory is more strongly correlated with IQ than traditional measures of (declarative) working memory. Hence, a more detailed understanding of how different aspects of working memory performance relate to real world measures in children is of great practical relevance.

Finally, the study of children's working memory for procedural representations such as task rules also has important theoretical implications. Specifically, the distinction between procedural and declarative memory has traditionally been considered in the context of long-term memory. In this context procedural ("implicit") memory refers to memory for skills, and declarative memory refers to memory for facts and events. In contrast, theories of short-term memory and working memory do not typically distinguish between the representation of declarative and procedural information. An exception to this is Oberauer's (2009) model of working memory, which does incorporate such a distinction: According to this theory, procedural and declarative working memory are separate but parallel subsystems. Whereas the former makes (declarative) representations available for processing; the latter enables processing through the representation and selection of procedures. The assumption that declarative and procedural working memory are separate subsystems with distinct capacity limits has been investigated by two adult studies, with mixed results (Barrouillet, Corbin, Dagry & Camos, 2015; Gade, Druey, Souza & Oberauer, 2014). Both of these studies employed a paradigm in which procedural and declarative working memory load was varied orthogonally. If procedural and declarative working memory have separate capacity limits, then increasing procedural working memory load should not affect declarative working memory, and vice versa. Using this method, Gade et al. (2014) found no substantial cross-task load effects, consistent with Oberauer's (2009) independence hypothesis. In contrast, Barrouillet et al. (2015) found that adults' performance on a (declarative) complex span task was affected by the procedural complexity of the concurrent processing task. Hence, their results argue against Oberauer's independence hypothesis.

Given these conflicting results, it is not yet clear whether the storage and manipulation of declarative and procedural representations in working memory are governed by separate subsystems or not. However, regardless of whether procedural and declarative working memory are distinct or not, in order to fully understand how working memory capacity affects the flexible control of behavior, working memory research must focus on the representation of procedural information alongside the representation of declarative information. For the practical and theoretical reasons highlighted above, the three experiments reported here focused on how children's ability to maintain and implement procedural representations develops throughout childhood (Experiments 1&2), and on the mechanisms responsible for this developmental change (Experiment 3).

To our knowledge, there are only two previous developmental studies which have investigated children's working memory capacity for task rules. To measure working memory capacity for task rules as a function of age, both of these studies used the set-size effect described Hick's law. Given that the RT set-size effect is considered an index of working memory capacity in the adult literature (see above), and given that working memory capacity increases throughout childhood (e.g., Gathercole, 1999), the set-size effect described by Hick's law should capture age differences in children's working memory capacity for task rules. Firstly, Davidson, Amso, Anderson and Diamond (2006) conducted a large-scale study investigating working memory, inhibition and cognitive flexibility in children (aged 4 to 13) and young adults. As part of a computerized battery, children completed a choice RT task in which either two (low working memory load) or six abstract shapes (high working memory load) were mapped onto one of two responses. The authors predicted that there should not be an interaction between working memory load and age. Instead, they assumed it should be harder to hold more rules in mind for everyone, regardless of age, though it is not clear what the foundation is of this prediction, given that a large body of research has shown working memory capacity to increase with age. Davidson et al.'s (2006) results are further complicated by an apparent speed-accuracy trade-off in their data for the ages of 6 through to young adulthood: Whereas the accuracy set-size effect decreased with age, the RT set-size effect increased with age. A subsequent study by Kiselev, Espy and Sheffield (2009) required children aged 4 to 6 to complete a choice RT task in which either 2 stimuli (low working memory load) or 4 stimuli (high working memory load) mapped onto one of two responses. In contrast to Davidson et al.'s (2006) findings, the results showed that the RT set-size effect decreased with age. Clearly, these two studies have yielded inconsistent results. Additionally, both of these studies confounded stimulus set-size with stimulus frequency. For example, in Kiselev et al.'s (2009) experiment participants completed 20 trials in the 2 S-R condition, and 20 trials in the 4 S-R condition. Because of this, each stimulus was presented twice as often in the 2 S-R condition. The number of stimulus repetitions was also greater in the 2 S-R condition. Both of these factors (stimulus frequency and immediate stimulus repetitions) are known to be inversely related to RT (Hyman, 1953; Van ‘t Wout, 2018), and it is therefore possible that faster responses in the 2 S-R condition were caused (at least in part) by these factors. Indeed, Van ’t Wout (2018) has shown that the set-size effect seen in adults is reduced by two-thirds when stimulus-frequency and recency are matched between set-size conditions. For this reason, the experiments reported here manipulated set-size, whilst keeping stimulus frequency between working memory load conditions constant. This was achieved in the following way: To start with, all children practiced a stimulus identification task with 5 S-R mappings, until 90% of all responses were accurate. Subsequently, the experimental session featured blocks of both high working memory load (featuring all 5 S-R mappings – ABCD&E) and lower working memory load (featuring only 3 or 4 out of 5 S-R mappings). By varying which specific S-R mappings featured in low working memory load blocks (e.g., ABC in one block, and BCD in another), it was ensured that overall, each S-R mapping occurred equally often in each set-size condition. Additionally, Experiment 3, which was designed to investigate the underlying causes of the age difference in working memory for task rules, avoided this confound altogether by using a paradigm in which the exact same stimuli and responses were used in the low working memory load and high working memory load conditions.

In summary, little remains known about how children’s ability to maintain and act on task rules held in working memory develops throughout childhood. The two previous choice RT experiments that have investigated the interaction between age and working memory load have found conflicting results. Furthermore, these studies confounded number of S-R rules with stimulus frequency. For this reason, Experiments 1 and 2 reported here measured working memory capacity by varying the number of stimulus-response rules in a RT experiment. Stimulus frequency was kept constant across set-size conditions, and the increase in RT with set-size was used as an index of working memory capacity. Experiment 1 examined whether children’s working memory capacity for task rules increases throughout childhood. Experiment 2 investigated (and dismissed) the possibility that age-related differences in the set-size effect are simply caused by overall RT differences between the age groups. Finally, Experiment 3 sought to determine whether or not the age difference in working memory capacity for task rules could be explained by age differences in strategic ability.

Experiment 1

Method

Participants

In total, 72 children from a local primary school in the UK took part in Experiment (see Table 1): 24 Year 1 children (aged between 5 years 8 months and 6 years 8 months, *M*=6:3), 24 Year 3 children (aged between 7 years 8 months and 8 years 7 months, *M*=8:1) and 24 Year 5 children (aged between 9 years 8 months and 10 years 8 months, *M*=10:1). Experiments 1, 2 and 3 were approved by the Faculty of Science Human Research Ethics Committee of the University of Bristol, and informed consent was obtained from parents prior to participation. The data of seven children (2 Year 1 children, 4 Year 3 children and 1 Year 5 child) were discarded and replaced as their mean RTs or error rates deviated from the year group mean by more than 3 standard deviations, or because they had neglected to respond to the central fixation on more than 50% of all trials.

			Age		Gender
		Mean	**Min**	**Max**	**F/M**
Experiment 1	**Year 1**	6:3	5:7	6:7	16/8
	Year 3	8:1	7:7	8:7	17/7
	Year 5	10:1	9:7	10:7	13/11
Experiment 2	**Year 4**	9:4	8:8	9:8	14/10

Table 1 Summary of participant details (age in years:months, and gender) from Experiments 1 and 2.

Apparatus

Participants were tested individually, in a quiet space outside of the classroom. Children were seated behind an Elo touchscreen monitor, which was connected to a Toshiba laptop. The experiment was programmed in Psychopy (Peirce, 2007), and the data were pre-processed and analyzed using excel and SPSS, respectively.

Design

The experimental task required children to respond to one of five different aliens (purchased from Etsy), each of which was characterized by a unique coloring and appearance. Each alien (approximate visual angle 4°) was associated with a unique target location on the computer screen (see Figure 1). Five identical spaceships (approximate visual angle 4°) were presented at the target locations (Figure 1), which were equally spaced in an arc of a fixed radius from the fixation point. The visual angle from the fixation point to the center of each spaceship was approximately 16°. On each trial**,** the child was instructed to respond to a centrally presented alien by touching its corresponding spaceship. Set-size was manipulated by varying the number of S-R mappings (aliens and their corresponding spaceships) in a block: Each block featured either 3, 4 or 5 of the aliens and their corresponding spaceships.

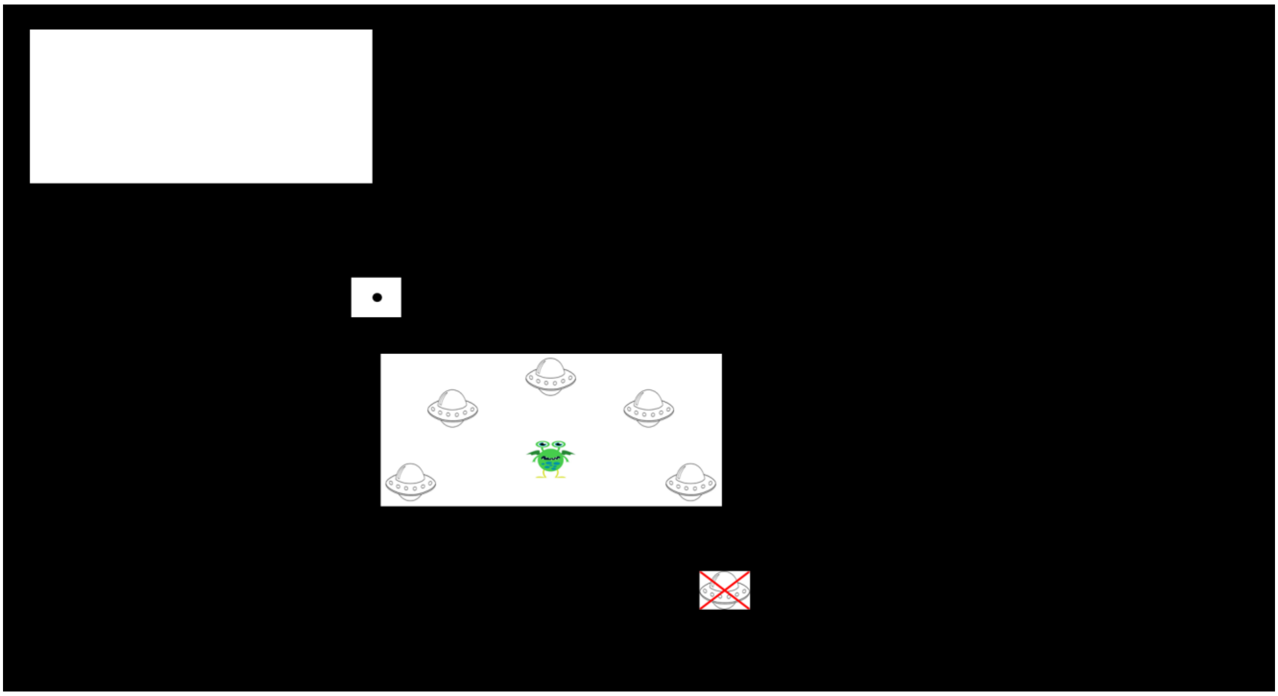

Figure 1 Example of a trial sequence in Experiment 1.

At the start of each block, participants were shown which 3, 4 or 5 aliens would feature in that block. Each trial began with a 500ms blank, followed by a fixation dot (duration of 1500ms). The fixation dot was succeeded by a centrally presented target alien, and the spaceships associated with that block (see Figure 1). Participants were instructed to touch the fixation dot, and then the correct spaceship, as quickly as possible. After a correct response, the alien would appear in its spaceship (duration 1000ms); after an incorrect response, a spaceship with a red cross would appear in a central location. Because stimulus repetitions are known to inflate the set-size effect (Hyman, 1953), the same alien never repeated from one trial to the next.

In total each child completed 15 blocks of 12 trials each: 5 blocks with 5 aliens, 5 blocks with 4 aliens, and 5 blocks with 3 aliens. Different (combinations of) aliens featured in each of the 3 and 4 alien blocks, in order to ensure that, across the entire experiment, each of the 5 aliens occurred equally often in each of the three set-size conditions. The trial procedure was identical across the three set-size conditions, with one exception: Only the spaceships that were possible response options in a particular block appeared on screen. This was done to avoid confusion (over which spaceships were possible response options), as the configuration of aliens in a block changed frequently (every 12 trials). The order of set-size conditions was counterbalanced between participants, as was the assignment of aliens to target locations. At the end of each block, an alien appeared with a speech bubble saying “Well done!”.

Procedure

At the start of the experiment, all children were told that they were about to play a computer game. It was explained that, in this game, five aliens were lost in space, and that the child had to help the aliens find their way back home. The child was then shown two consecutive instruction screens: The first screen showed all five aliens (each presented at its respective response location); the second screen showed only the five space ships. At this stage the experimenter pointed at each of the five spaceships in turn, and asked the child to recall and describe the alien associated with the spaceship (e.g., “the blue one”, “the red one with the pointy ears”).

Following these verbal instructions, each participant completed 3 separate tasks: First, all participants completed a baseline task of 20 trials. This baseline task served to (1) familiarize participants with the use of the touch screen computer, and (2) obtain a baseline measure of RT for all participants. Each trial began with a blank (500ms), followed by a central fixation dot (1500ms), after which a spaceship would appear in one of the five response locations. Participants were instructed to touch the fixation dot, and then the spaceship, as quickly as possible. Each individual’s mean RT obtained with this task could then be used in the analysis to account for age-related differences in overall RT.

Second, all children then practiced the experimental task, featuring all five aliens. The aim of this practice session was to familiarize participants with the experimental task and to ensure that, through practice, S-R mappings would be implemented as a “prepared reflex” (Hommel, 1998; also see Barrouillet et al., 2015), no longer reliant on a declarative representation of the instructions. This practice session consisted of a maximum of 10 blocks of 10 trials, with each alien appearing twice in every block. The practice was terminated once participants achieved 90% accuracy in a practice block. The practice session was immediately followed by the experimental session. In total, the experiment lasted approximately 15-20 minutes.

Results

Prior to analyses, very long (> 5000ms) reaction times (2.0% of all trials) were removed from the data set. Trials on which the participant had failed to respond to the central fixation dot (7.4% of all trials) were also excluded from the analysis. The remaining data were pooled across blocks for each set-size condition. Throughout the manuscript, the number following the ± symbol indicates the standard error of the mean (SEM).

The effect of age on the set-size effect

Reaction times

A mixed-design ANOVA with the within subject factor set-size (3, 4 or 5 S-R rules) and the between subjects factor year group (Year 1, Year 3 or Year 5) was run on the mean RT data (see Table 2 & Figure 2). This analysis showed that RTs decreased significantly with age, $F(2,69)=31.21$, $p<.001$, $\eta_p^2=.475$. The decrease in RT was much bigger from Year 1 to Year 3 (349 ± 65 ms, $F(1,46) =31.01$, $p<.001$, $\eta_p^2=.403$), than from Year 3 to Year 5 (76 ± 46 ms, $F(1,46) =1.96$, $p=.168$, $\eta_p^2=.041$).

RTs also increased significantly as a function of set-size, $F(2,138)=29.66$, $p<.001$, $\eta_p^2=.301$: The mean set-size effect (measured as the difference between the 3 and 5 S-R condition) across age groups was 97 ± 15 ms. Additionally, this set-size effect decreased with age, $F(4,138)=3.92$, $p=.006$, $\eta_p^2 =.102$ (the interaction between year group and the linear component of set-size was also significant, $F(2,69)=6.18$, $p=.003$, $\eta_p^2 =.152$): The set-size effect was 166 ± 36 ms in Year 1, 75 ± 17 ms in Year 3 and 49 ± 17 ms in Year 5. The decrease in set-size effect was much bigger from Year 1 to Year 3 (92 ± 42 ms, $F(1,46) =5.33$, $p=.025$, $\eta_p^2 =.104$), than from Year 3 to Year 5 (26 ± 18 ms, $F(1,46) =1.22$, $p=.276$, $\eta_p^2=.026$).

Error rates

The error rates were analyzed using the same 3x3 mixed design ANOVA described above (see Table 2 & Figure 2). The only significant effect in this analysis was the main effect of age: The mean error rate in Year 1 (9.2 ± 1.2%) was higher than in Year 3 (4.2 ± 0.6%) and Year 5 (3.0 ± 0.4%), $F(2,69)=16.74$, $p<.001$, $\eta_p^2=.327$. As in the RT data, the difference in error rates between Year 1 and the other two age groups was significant (with Year 3 $F(1,46)=14.38$, $p<.001$, $\eta_p^2=.238$; with Year 5 $F(1,46)=25.25$, $p<.001$, $\eta_p^2=.354$); but the difference between Year 3 and Year 5 error rates was not significant, $F(1,46)=2.26$, $p=.140$, $\eta_p^2=.047$. There was no significant main effect of set-size (set-size effect of -0.2 ± 0.5%, $F(2,138)=.860$, $p=.426$,

η_p^2=.012), and the set-size effect did not differ significantly between Year 1 (0.0 ± 0.6%), Year 3 (-0.5 ± 0.4%) and Year 5 (-0.2 ± 0.2%), F(4,138)=1.05, p=.355, η_p^2=.030.

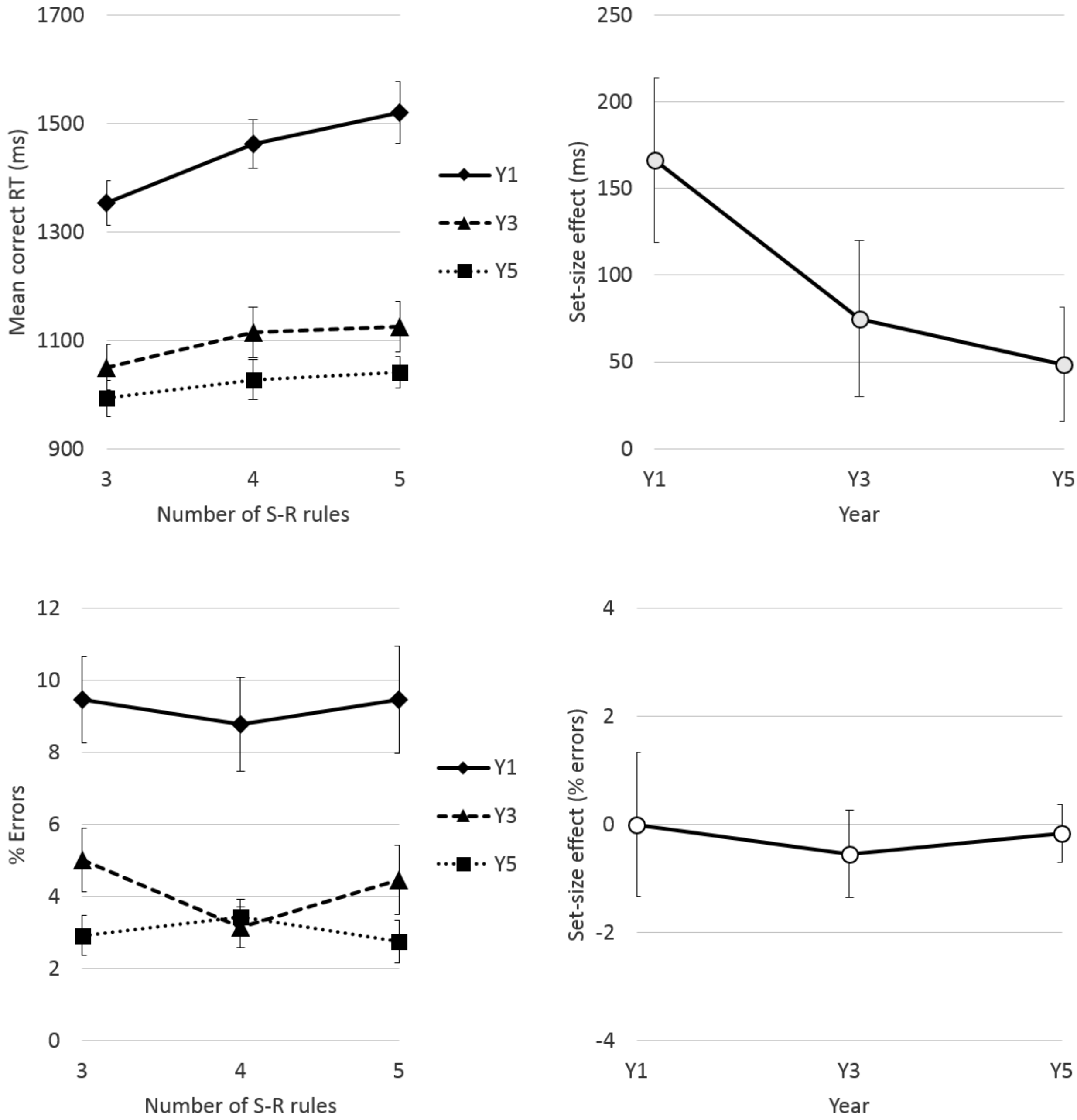

Figure 2 Left: Mean RT ± SEM (top) and % error ± SEM (bottom) data plotted separately for each year group, as a function of the number of S-R rules. Right: The same data plotted as difference scores (between the 5 S-R and 3 S-R condition) or "set-size effects", as a function of year group.

		Overall RT	Experimental session 3 S-R	4 S-R	5 S-R	Set-size effects Absolute	Proportional
Mean RT	**Y1**	1445	1354	1462	1520	166	12.3
	Y3	1096	1050	1114	1125	75	7.4
	Y4	1965	1916	1964	2016	100	5.7
	Y5	1021	993	1028	1041	49	5.6
Errors	**Y1**	9.2	9.5	8.8	9.5	0.0	
	Y3	4.2	5.0	3.2	4.5	-0.5	
	Y4	5.0	4.7	5.6	4.6	-0.1	
	Y5	3.0	2.9	3.4	2.8	-0.2	

Table 2 Mean correct RTs (ms) and % error rates in Experiments 1 (Years 1, 3 & 5) and Experiment 2 (Year 4).

Proportional analysis

The above reported analysis of reaction times found the set-size effect to decrease significantly with age, suggesting that working memory capacity for task rules increases with age. However, because there was also a substantial difference in overall RT between the three year groups, it is possible that the larger Year 1 set-size effect is in fact proportionally similar to the Year 3 and Year 5 set-size effects, once overall differences in RT are taken into account. Indeed, mean correct RTs in the baseline condition, in which participants were simply required to touch a spaceship that appeared in one of the five target locations, decreased significantly as a function of age, F(2,96)=5.83, p=.005, η_p^2=.145: Baseline RTs were slowest for Year 1 children (716 ± 17 ms), and slower for Year 3 (646 ± 16 ms) than for Year 5 children (633 ± 22 ms). The mean baseline RT difference between Year 1 and Year 3 children was significant, F(1,46)=8.74, p=.005, η_p^2=.160, whereas the mean baseline RT difference between Year 3 and Year 5 children was not, F(1,46)=.27, p=.608, η_p^2=.006.

To investigate the possibility that the above observed age differences in the RT set-size effect were simply a reflection of overall age differences in RT, the experimental data were scaled using the mean RT from the baseline condition. For each individual, the experimental data were scaled by subtracting the baseline RT from the mean RT in each set-size condition, and dividing this by the baseline RT/100 [(MeanRT-BaselineRT)/(BaselineRT/100)]. Figure 3 clearly shows that, even once the experimental data are scaled in this way to account for overall differences in RT, the set-size effect remains larger for the Year 1 children than for the children from Years 3 and 5 (set-size effects of 24%, 12% and 8%, respectively), $F(4,138)=3.36$, $p=.012$, $\eta_p^2=.089$. The interaction between year group and the linear component of set-size was also significant using the scaled RT data, $F(2,69)=5.29$, $p=.007$, $\eta_p^2=.133$. This proportional analysis therefore suggests that the observed effect of age on the set-size effect is not caused (exclusively) by age differences in overall RT, and instead reflects a fundamental difference in working memory capacity.

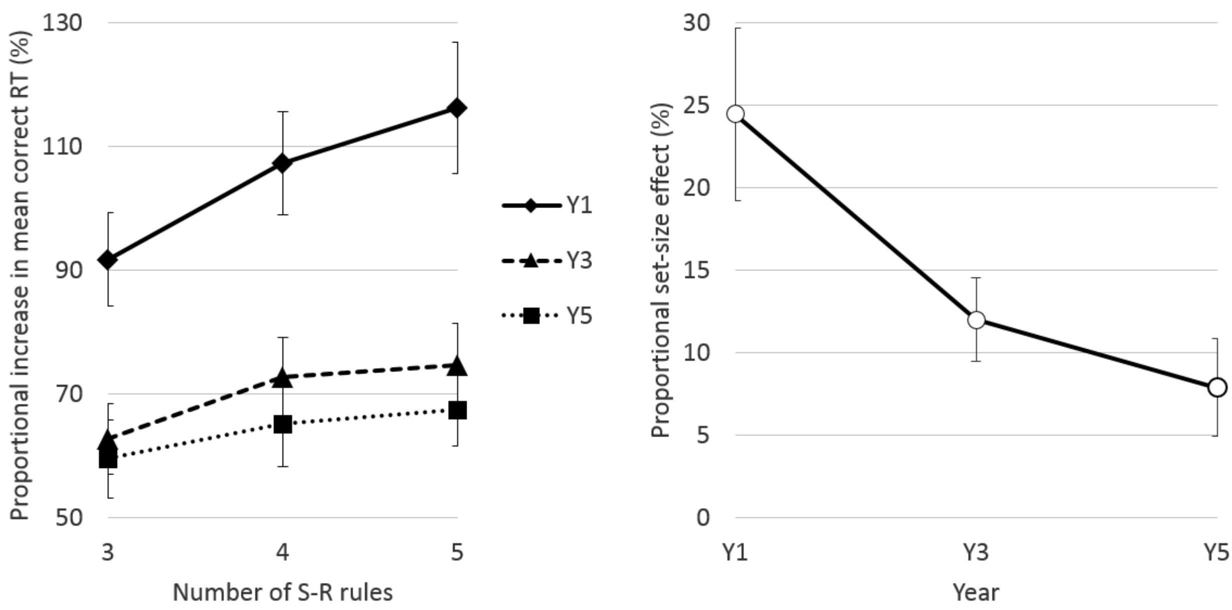

Figure 3 Mean correct RT (± SEM) data plotted as a proportional increase from baseline. Left: Proportional scores plotted separately for each year group, as a function of the number of S-R rules. Right: The same data plotted as difference scores (between the 5 S-R and 3 S-R condition) or "set-size effects", as a function of year group.

RT distribution analysis

To further explore the effect of age on the RT set-size effect, RT distributions were examined. RT distributions can be informative as they reveal whether the reported mean RT differences are representative of the entire RT distribution. For this analysis, cumulative distribution functions (CDFs) were computed by rank ordering the correct RTs (for each condition and each participant separately), and computing deciles (10th-90th percentile). These CDFs are shown in Figure 4, on the left. In the interest of clarity, the same data have also been captured in terms of “fast” (averaged over the 10th-40th percentiles) and “slow” (averaged over the 60th-90th percentiles) responses (on the right).

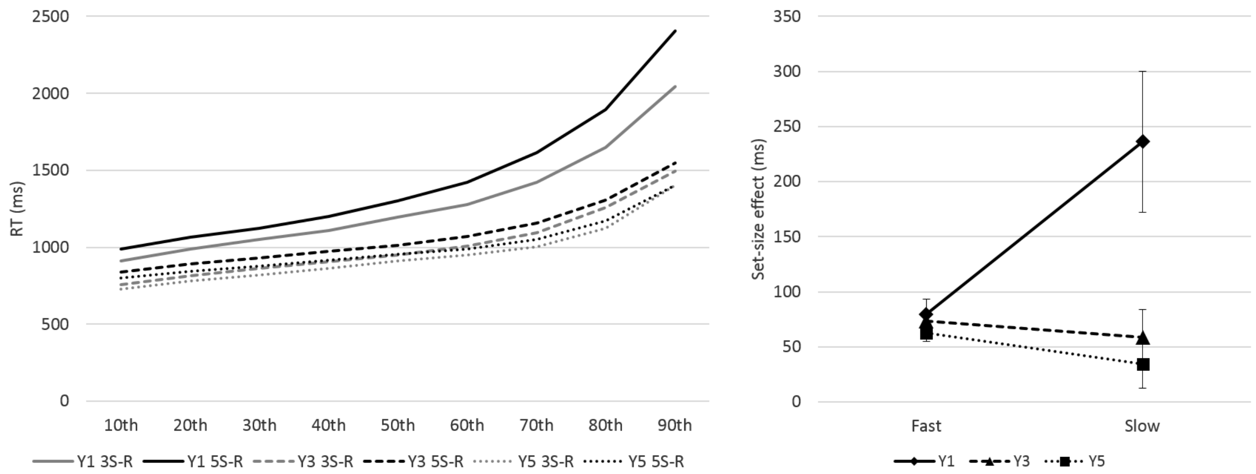

Figure 4 Left: Cumulative distribution functions in the 3 S-R and 5-R condition, plotted separately for each year group. Right: Set-size effects (in ms ± SEM) for “fast (10^{th}-40^{th}) and “slow” (60^{th}-90^{th}) responses, plotted separately for each year group.

A 2 (fast/slow) x 3 (number of S-R rules) x 3 (year group) mixed ANOVA was run on these data. The key result here is the 3-way interaction between year group, set-size and fast/slow, $F(2,69)=7.73$, $p=.001$, $\eta_p^2=.183$. As can be seen from Figure 4, this interaction reflects that the set-size effect is much larger for slower than for fast responses, but only for the Year 1 children. This suggests that younger children do struggle more in the 5 S-R condition, but only on a proportion of trials (i.e., the slow trials, $F(2,69)=6.98$, $p=.002$, η_p^2 =.114). Indeed, at the fast end of the distribution, there is no significant difference in the set-size effect between the age groups (F(2,69)=.713, p=.494, η_p^2=.020), suggesting that on a substantial proportion of trials, young children's working memory performance is comparable to that of older children.

Summary

Experiment 1 investigated whether children's ability to maintain a set of task rules in a heightened state of accessibility improves with age. To this end, children from 3 different age groups (Year 1, Year 3 and Year 5) performed a computerized task involving (subsets of) 5 arbitrary stimulus-response rules. As an index of working memory capacity, Experiment 1 used the increase in RT typically observed with an increase in set-size (also known as "Hick's law").

The results showed that the set-size effect was much larger in Year 1 children than in children in Years 3 and 5. However, given that the mean RT was also much higher in Year 1 children, it seemed possible that the larger set-size effect found for Year 1 participants actually reflected proportional scaling, and not a genuine difference in working memory capacity. To assess this possibility, the experimental data were coded as a proportional increase from baseline performance. This "proportional RT set-size effect" was still much larger for children in Year 1 than for those in Years 3 and 5, suggesting that the increase in the set size effect with age cannot be explained by a proportional scaling account. Even so, because the proportional analysis was conducted post-hoc, a second experiment was specifically designed to test directly the contribution of overall age differences in RT to the increase in set-size effect with age.

Experiment 2

Aim

To investigate directly whether the proportional scaling hypothesis could account for the results of Experiment 1, data were collected from another group of older (Year 4) children, using a more difficult version of the experimental task. The difficulty of the task was increased with the aim of reducing the overall RT difference between this new group of older children and the youngest (Year 1) children from Experiment 1: If the interaction between age and set size found in Experiment 1 reflects proportional scaling, then this new sample of older (Year 4) children should also show a larger set-size effect than the Year 3 or Year 5 children from Experiment 1 when their overall RT corresponds to that of the Year 1 children from Experiment 1.

To enable the comparison of set-size effects between Experiments 1 and 2, it was essential to increase only the difficulty of the task (e.g., by increasing the difficulty of the visual discrimination), and not its procedural complexity (e.g., by increasing the number or complexity of the task rules). For the same reason, it was important for this new task to be as similar as possible to the task used in Experiment 1.

Method

Participants

Twenty-four Year 4 children (aged between 8 years 10 months and 9 years 10 months, *M*=9:4) took part in Experiment 2. Using the same criteria as in Experiment 1, the data of 4 children were discarded and replaced.

Design and procedure

For the reasons mentioned above, the task used in Experiment 2 was identical to the task used in Experiment 1, apart from two modifications: 1) The aliens were presented in grayscale rather than color, so that it would be harder to discriminate one alien from another; 2) The target alien was presented amongst two distractor aliens. These distractor aliens were selected randomly on each trial from a pool of 4 distractor aliens (not part of the target set). On each trial, the target alien and two distractors were presented in one of three randomly allocated positions each (left, middle, center; see Figure 5). The target alien was always presented upside-down (180°), whereas the distractors were presented at varying angles

(0°, 90° or 270°). Hence, prior to selecting a response, the participant had to identify the target alien amongst distractors.

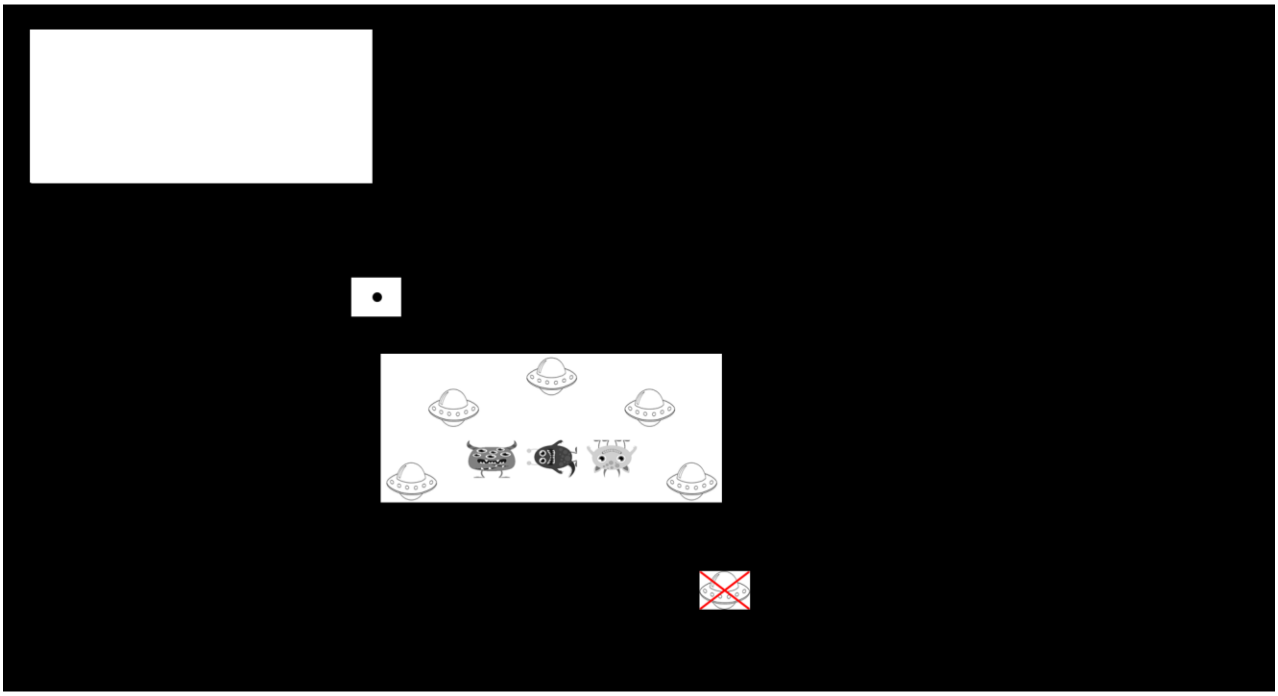

Figure 5 Example of a trial sequence in Experiment 2

As in Experiment 1, participants first performed the baseline task, followed by a practice session. In this practice session participants were presented with a single (grayscale) alien (i.e., no distractors were present); they performed a maximum of 10 consecutive blocks of this task until 90% accuracy was achieved. This was followed by a single practice block of 12 trials during which the target alien was presented upside-down, amongst distractors. As in Experiment 1, the experimental session consisted of 15 blocks (5 in each set-size condition). In total, the experiment lasted approximately 20-25 minutes.

Results

As in Experiment 1, very long (>5000ms) reaction times (2.0% of all trials), and responses following a missed fixation (6.3% of all trials) were excluded from the analysis. Below is a brief description of the Year 4 data in isolation, followed by a more detailed comparison with the Experiment 1 data.

Year 4 reaction times and error rates

The modifications made to the experimental task were successful in making the task much more difficult to perform: The mean RT for the Year 4 children was 1965 ± 79 ms (compared to 1445 ± 450 ms for the Year 1 children in Experiment 1). There was a significant set-size effect of 100 ± 26 ms, $F(1,23)=15.34$, $p=.001$, $\eta_p^2=.400$ (linear trend). The mean error rate was 5.0 ± 0.6%, and there was no set-size effect (0.0 ± 0.4%) in the error data.

Cross-experiment comparison: The effect of age on the set-size effect

Reaction Times

To compare the data from the Year 4 children with the data obtained in Experiment 1, a 3 (number of S-R rules) x 4 (year groups) mixed-design ANOVA was run. As in Experiment 1, the analysis found a significant interaction between year group and the set-size effect, $F(3,92)=4.07$, $p=.009$, $\eta_p^2=.117$ (see Table 2 & Figure 6).

The most important comparison is between the Year 1 and the Year 4 children's data: The mean RT was significantly larger (by 520 ms) for the Year 4 children than for the Year 1 children, $F(1,46)=32.98$, $p<.001$, $\eta_p^2=.418$. Despite this, the set-size effect remained larger for the Year 1 children (166 ± 36 ms) than for the Year 4 children (100 ± 26 ms), $F(1,46)=2.28$, $p=.138$, $\eta_p^2=.047$. Although this difference in set-size effects was not significant, this pattern of results still counts against the proportional scaling hypothesis as the *only* explanation of the Experiment 1 data, as this hypothesis would predict a larger set-size effect for the Year 4 children.

Error rates

For the error data, the same 3 (number of S-R rules) x 4 (year groups) mixed-design ANOVA only showed a significant main effect of year group, $F(3,92)=12.55$, $p<.001$, $\eta_p^2=.290$; reflecting a higher error rate for the Year 1 children (9.2 ± 1.2%) compared to the Year 3 (4.2 ± 0.6%), Year 4 (5.0 ± 0.6%) and Year 5 (3.0 ± 0.4%) children. The main effect of set-size and the interaction between set-size and year group were not significant, $F(2,184)=.194$, $p=.824$, $\eta_p^2=.002$ and $F(6,184)=1.187$, $p=.315$, $\eta_p^2=.037$, respectively.

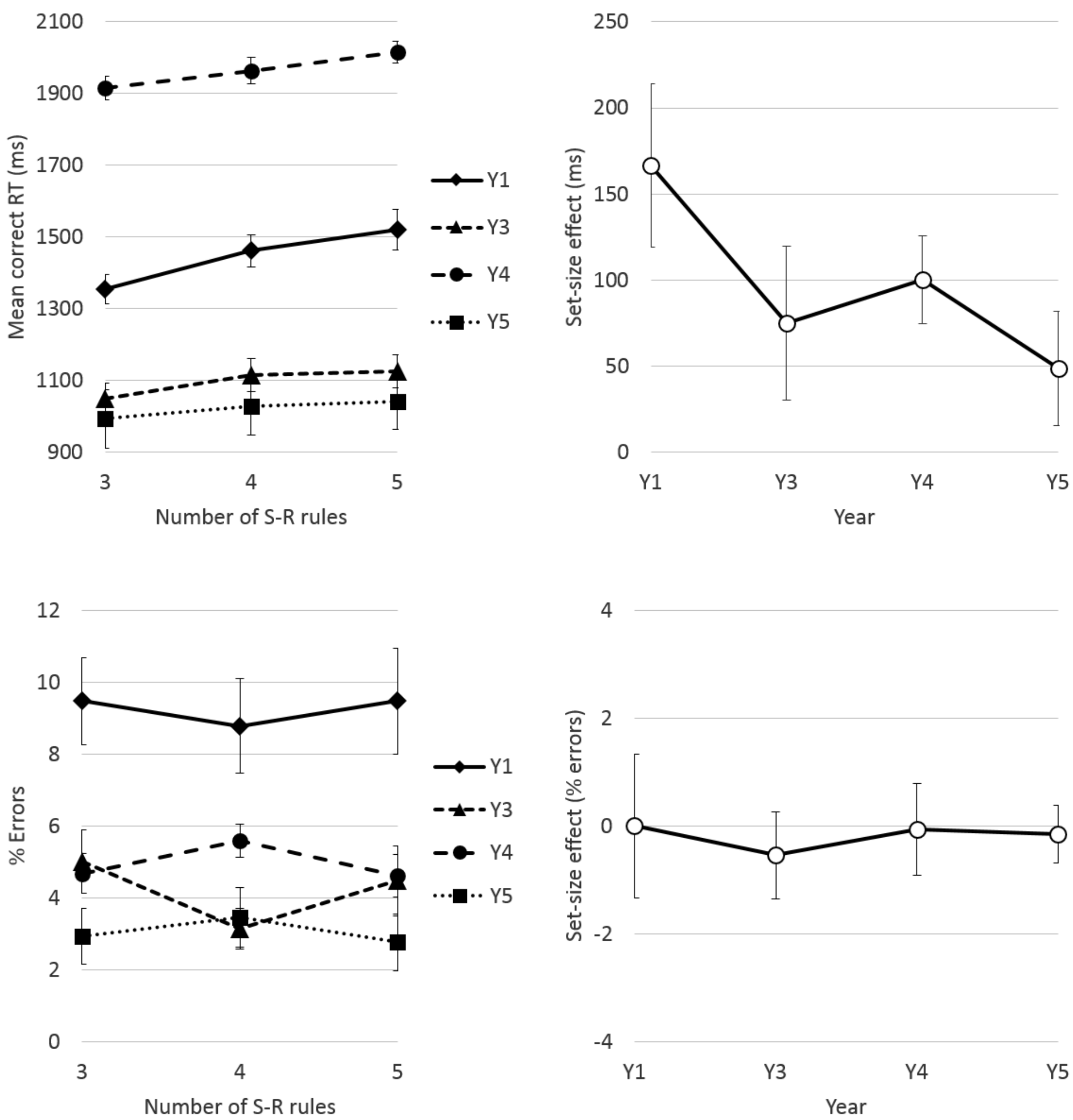

Figure 6 Left: Mean RT ± SEM (top) and % error ± SEM (bottom) data plotted separately for each year group, as a function of the number of S-R rules. Right: The same data plotted as difference scores (between the 5 S-R and 3 S-R condition) or "set-size effects", as a function of year group.

Proportional analysis

As in Experiment 1, the scaled data were also analyzed. Because it was no longer meaningful to scale the data to the baseline condition (as only the difficulty of the experimental session had been increased in Experiment 2, whereas the baseline condition had remained the same), the data were now scaled to the easiest (3 S-R) condition (see Figure 7).

This proportional analysis found that the scaled Year 4 set-size effect of 5.7 ± 1.4% was very similar to the scaled Year 5 set-size effect (5.6 ± 1.6%), $F(1,46)<.01$, $p=.969$, $\eta_p^2 <.001$; and significantly smaller than the scaled Year 1 set-size effect of 12.3 ± 2.6%, $F(1,46)=5.20$, $p=.027$, $\eta_p^2 =.102$.

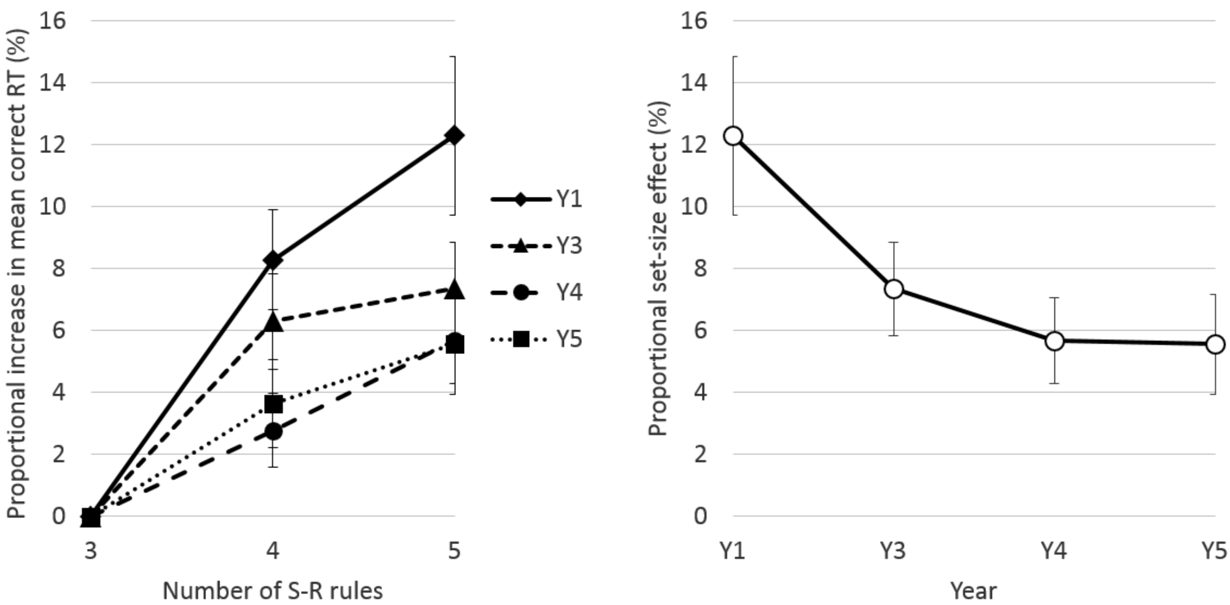

Figure 7 Mean correct RT ± SEM data plotted as a proportional increase from the 3 S-R condition. Left: Proportional scores plotted separately for each year group, as a function of the number of S-R rules. Right: The same data plotted as difference scores (between the 5 S-R and 3 S-R condition) or "set-size effects", as a function of year group.

Summary

The main purpose of Experiment 2 was to directly test the possibility that age-differences in the set-size effect are due to proportional scaling. If this was the case, then the Year 4 children (who had the largest mean RT by >500ms) should have shown the largest set-size effect. However, this was not the case. In absolute terms, the set-size effect remained smaller for the Year 4 children in Experiment 2 than for the Year 1 children in Experiment 1 (though not significantly so). Once proportionally scaled, the Year 4 set-size effect was significantly smaller than the Year 1 scaled set-size effect, and appeared more similar to the Year 5 set-size effect in Experiment 1. These findings argue against a proportional explanation of the effect of age on the RT set-size effect, and provide further support for fundamental age-related differences in children's working memory capacity for task rules.

Experiment 3

Aim

Experiments 1 and 2 established that children's ability to maintain stimulus-response rules in working memory increases with age. Proportional analyses, together with the results of Experiment 2, confirmed that the age-related difference in RT set-size effect was not caused by overall age differences in speed of responding. Experiment 3 aimed to build on these findings in two specific ways: Firstly, by replicating this result using a different paradigm which circumvents a particular experimental confound (described below), and secondly, by further investigating the underlying cause of the observed difference in working memory capacity for task rules.

With regards to this first aim, as noted in the Introduction, in set-size experiments it is vital to control for stimulus frequency, as the resulting inverse correlation between set-size and stimulus frequency is known to inflate set-size effects (Van 't Wout, 2018). Although Experiments 1 and 2 equated stimulus frequency for each set-size condition across the experimental session, within each block of 12 trials, this inverse relationship remained intact[1]. To address this potential confound, Experiment 3 manipulated working memory load in a different manner. Specifically, this experiment required participants to classify 8 stimuli using either a set of 8 arbitrary S-R rules (high working memory load condition), or 2 categorical rules (low load working memory load condition). Accordingly, in this particular paradigm (described in more detail below), memory load could be varied whilst keeping the number of stimuli and responses constant between the high and low load conditions.

The second aim of Experiment 3 was to investigate the cause(s) of the observed age difference in working memory capacity for task rules. There are two dominant accounts of developmental differences in working memory performance. The first ("capacity account") assumes that working memory capacity has a structural limit, which increases with age (e.g., Cowan, 2001). The second ("efficiency account") assumes that the available capacity of working memory does not increase with age, but instead argues that older children carry out the component processes that support working memory more efficiently, freeing up relatively more resources for storage (Case et al., 1982). One way in which these age-related increases in efficiency might manifest is through the use of more efficient maintenance mechanisms (Hulme, Thompson, Muir, & Lawrence, 1984).

Indeed, there is some evidence to suggest that the age difference in working memory performance could be caused by age differences in strategic ability (e.g., Cowan, Cartwright, Winterowd & Sherk, 1987; Tam et al., 2010). For example, Tam, Jarrold, Sabatos-DeVito, & Baddeley (2010) found that the detrimental effect of verbal processing on recall was greater for 8-year olds compared to 6-year olds. However, others have shown that the age difference in working memory capacity remains even when rehearsal is prevented through articulatory suppression (Cowan et al., 2011). Hence, studies investigating whether an increase in the use of verbal strategies might underly the age difference in working memory capacity have yielded mixed results. Furthermore, most of these studies have focused on people's ability to maintain declarative representations (see Camos, Lagner, & Barrouillet, 2009; Hudjetz & Oberauer, 2007; Tam et al., 2010). It is

[1] For example, in a block with 5 aliens, each alien would occur on average 2.4 times; whereas in a block with 3 aliens, each alien would occur on average 4 times.

not yet known whether those same mechanisms can also be used to support the active maintenance and implementation of procedural representations in working memory. One intuition is that verbal rehearsal, thought to be a powerful tool for the maintenance of declarative information, will be less effective for maintaining procedural information in working memory (because the latter might not be stored in a format suitable for articulatory rehearsal). Consistent with this intuition are the results from Van 't Wout, Lavric, and Monsell (2013), who have shown that adults' execution of well-practiced S-R mappings is not influenced by phonological similarity, suggesting that adults do not use sub-vocal rehearsal to maintain procedural representations. Based on the evidence reviewed above, it remains unclear whether age difference in strategic (specifically verbal) ability could explain the observed developmental difference in working memory performance. To test this possibility, Experiment 3 measured working memory capacity for task rules whilst children performed a verbal secondary task (articulatory suppression), a non-verbal secondary task (foot tapping) or no secondary task (control condition). If the age difference in working memory capacity for task rules is caused by a basic difference in capacity, then the effect of the secondary tasks on memory load should not differ across age groups. Alternatively, it is possible that the age difference in working memory capacity is strategic, either because younger children do not use such strategies (a production deficiency), or because they do not benefit from using such strategies to the same extent that older children do (a utilization deficiency; e.g., Miller, Seier, Barron & Probert, 1994). Either way, if the age difference in working memory capacity is strategic, then the effect of age on memory load should be reduced or eliminated when children have to perform a secondary task that blocks the use of that strategy. In the latter case, the nature of the secondary task (verbal or non-verbal) will therefore tell us something about the underlying maintenance mechanism. Specifically, if more efficient use of verbal strategies is at the root of older children's increased working memory capacity, then the age difference in working memory capacity should be attenuated by articulatory suppression, but not under foot tapping.

Method

Participants

Forty-four children from a local primary school took part in Experiment 3: 24 “older” children from Years 4 and 5 (aged between 8 years 10 months and 10 years 9 months, *M*=9:8) and 20 “younger” children from Year 1 (aged between 5 years 3 months and 6 years 3 months, *M*=5:9). One child from Year 1 did not compete the session, and the data from this child were excluded from the analysis.

Stimulus Set 1				Stimulus Set 2			
Low load		High load		Low load		High load	
L (edible)	R(inedible)	L	R	L (big)	R (small)	L	R
Cake	Book	Cake	Book	Horse	Rabbit	Horse	Rabbit
Bread	Pen	Pen	Bread	Cow	Cat	Cat	Cow
Apple	Hat	Apple	Hat	Elephant	Squirrel	Elephant	Squirrel
Cheese	Shoe	Shoe	Cheese	Giraffe	Bee	Bee	Giraffe

Baseline Stimuli	
L (green)	R (blue)
Bicycle	Scissors
Flag	Sun
Hand	Tree
Kite	Watch

Table 3 Tasks and picture names of object stimuli used in Experiment 3.

Design and procedure

In this experiment, children were required to respond to a set of 8 pictures of animals and objects with one of two response keys (see Table 3). The pictures were selected from the International Picture Naming

Project, Szekely et al., 2004) and were matched for naming latency (M = 829ms) and age of acquisition (AoA between age 1.98-3.79; Kuperman, Stadthagen-Gonzalez & Brysbaert, 2012). Each child performed a "high working memory load" condition and a "low working memory load" condition. In the low load condition, the organization of the eight S-R mappings could be described by a category rule: Specifically, if you can eat it, press left; if you can't press right (Stimulus Set 1). Or (Stimulus Set 2): If the animal is bigger than you, press left, if it is smaller than you, press right. For each stimulus set, in the high load condition, the eight stimuli were assigned arbitrarily to the two responses. Because of this, the low load condition required children to remember only two rules; whereas in the high load condition, children had to remember eight S-R rules. To measure the effect of working memory load within participants, two different stimulus sets were used (see Table 3), in order to avoid carry over effects from previous S-R associations: Half of the children performed the low load condition with Stimulus Set 1 and the high load condition with Stimulus Set 2, and vice versa.

To examine whether the previously found effect of age on working memory load is caused by older children's more efficient use of maintenance mechanisms, performance in the low and high load conditions was measured under three conditions: (1) control condition (no secondary task); articulatory suppression (AS; verbal secondary task); (3) foot tapping (FT; non-verbal secondary task). Both foot tapping (with one foot only) and articulatory suppression (saying "tick, tick, tick") were performed to the beat of a metronome set to 90 beats per minute. Children were instructed to ignore the metronome during the control (no secondary task) condition.

Each child performed 168 trials per working memory load condition (48 trials for each secondary task condition, plus 24 practice trials). In the addition to the high and low working memory load conditions, a baseline condition was also included to assess the effect of the secondary tasks on performance in the absence of any working memory load. In this baseline condition participants performed a color matching task: eight pictures (four green, four blue; see Table 3) had to be classified according to color. Two colored boxes were present on the left and right bottom of the screen in order to eliminate any memory load. All children performed 72 trials of the baseline condition first (24 trials for each of the three secondary task conditions), followed by seven blocks of each working memory load condition (high or low), consisting of one practice block plus two blocks for each of the secondary task conditions.[2] The order of working memory load (high or low) and the order of secondary task (tapping or AS) was varied between participants.

Results

Prior to analyses, very long (> 8000ms) and very short (<200ms) reaction times (1.0% of all trials) were removed from the data set. Note that a different cut-off for very long RTs was used in Experiment 3 (> 8000ms) compared to Experiment 1 (> 5000ms), because on average RTs were significantly greater in Experiment 3 (1377ms) than Experiment 1 (1210ms), $F(1,113)=5.610$, $p<.001$, $\eta_p^2=.333$. Importantly, the proportion of excluded responses was very similar in Experiments 1 (2.0%), 2 (2.0%) and 3 (1.0%) (Ratcliff, 1993).

The effect of a secondary task on performance in the absence of a working memory load

First, to examine whether tapping and AS were approximately equally difficult to perform in the absence of a memory load, mean correct RTs and errors in the baseline condition were analyzed[3]. For the mean correct RTs, a 2 (younger or older children) x 3 (control, AS or FT) ANOVA found a significant main effect of age, $F(1,41)=24.447$, $p<.001$, $\eta_p^2=.374$ (older children were 388ms faster than younger children), and a significant main effect of secondary task, $F(2,82)=10.142$, $p<.001$, $\eta_p^2=.198$, reflecting that children were faster in the absence of a secondary task (see Figure 8). However, a further 2 (younger or older children) x

[2] Note that in Experiment 3, unlike in Experiment 1, children were not trained until 90% accuracy was achieved. This is because unlike in Experiment 1, in Experiment 3 the low and high working memory load conditions involved different tasks, and so to train to 90% accuracy would have resulted in more training in one condition (high working memory load) than another (low working memory load), which would have introduced additional confounds.
[3] The first block of 12 trials (no secondary task) of the baseline condition was considered a practice block and was not included in the analysis.

2 (tapping or AS) ANOVA found no significant difference between tapping (1201±61ms) and AS (1140±53ms), F(1,41)=1.713 p=.198, η_p^2=.040. For the error data, the same 2 x 3 ANOVA revealed a significant main effect of age, F(1,41)=7.295, p=.010, η_p^2=.151, and a significant main effect of secondary task, F(2,82)=7.339, p=.001, η_p^2=.152 (control 3.3±1.0%, AS 7.4±1.1% and tapping 4.5±0.7%). A further 2 x 2 ANOVA showed that children made slightly more errors in the AS condition (7.4±1.1%) than in the tapping condition (4.5±.7%), F(1,41)=9.043 p=.004, η_p^2=.181. In both the RT and the error ANOVA, the interaction between secondary task and age was not significant, F's< 1.928. Hence, AS was a slightly more difficult secondary task than tapping, regardless of age.

The effect of a secondary task on the working memory load effect as a function of age

Mean correct RTs

To compare the effects of AS and FT in the high and low load conditions for the two different age groups, a 2 (load: high or low) x 3 (secondary task: control, tapping or AS) x 2 (age: younger or older children) ANOVA was run on the mean correct RT data.

This analysis found a significant main effect of load: overall, RTs were greater in the high load than in the low load condition, F(1,41)=20.463 p<.001, η_p^2=.333. Importantly, this effect of working memory load was much greater in the control condition (258±35ms), than under either tapping (105±44ms) or AS (113±48ms), F(2,82)=6.900, p=.002, η_p^2=.144 (two-way interaction). As can be seen from Figure 8, the reduced working memory load effect with a secondary task was caused mainly by slower RTs in the low load condition (rather than faster RTs in the high load condition). There was also a significant main effect of age, F(1,41)=25.234, p<.001, η_p^2=.381 (mean correct RTs were 1545±68ms and 1133±47ms for younger and older children, respectively); but there were no significant interactions with age (all F's < 1.898).

A separate 2 (working memory load: high versus low) x 2 (secondary task: FT versus AS) x 2 (age: younger or older children) excluding the control condition confirmed that the load effects did not differ significantly between the AS (113±48ms) and tapping conditions (105±44ms), $F(1,41)=.020$, $p=.887$, $\eta_p^2<.001$ (two-way interaction). The main effect of load was still significant, $F(1,41)=7.879$, $p=.008$, $\eta_p^2=.161$. There was no significant main effect of secondary task, $F(1,41)=1.126$, $p=.295$, $\eta_p^2=.027$: Children were approximately equally fast under AS (1336±56ms) or tapping (1369±56ms).

Errors

In the error data, a 2 (load: high versus low) x 3 (secondary task: control, tapping or AS) x 2 (age: younger or older children) found a significant main effect of load, $F(1,41)=17.772$, $p<.001$, $\eta_p^2=.302$, and a significant main effect of secondary task, $F(2,82)=3.638$, $p=.031$, $\eta_p^2=.081$: Children made more errors in the high load (12.7±1.7%) than in the low load condition (5.4±0.6%); and fewer errors in the control condition (8.3±0.9%) and under tapping (8.5±1.0%) than under AS (10.3±1.2%). Furthermore, the working memory load effect was larger for younger children (11.1±3.2%) than for older children (3.3±1.5%), $F(1,41)=5.171$, $p=.028$, $\eta_p^2=.112$; and the interaction between load and secondary task was also significant, $F(2,82)=6.128$, $p=.003$, $\eta_p^2=.130$.

A separate 2 (load: high versus low) x 2 (secondary task: tapping versus AS) excluding the control condition found that the working memory load effect was larger under articulatory suppression (8.7±2.3%) than under tapping (4.1±1.7), $F(1,41)=10.174$, $p=.003$, $\eta_p^2=.199$ (two-way interaction). However, because the working memory load effect was not significantly larger under AS compared to the control condition (7.4±1.8%), $F(1,41)=.869$, $p=.357$, $\eta_p^2=.021$ (two-way interaction); and because AS was a slightly more difficult secondary task even in the absence of a working memory load (see baseline analysis above); this result must be interpreted with caution. Finally, and most importantly, the 2 (load: high versus low) x 3 (secondary task: control, tapping or AS) x 2 (age: younger or older children) ANOVA found that the effect of the secondary task on the working memory load effect was not significantly modulated by age, $F(2,82)=1.669$, $p=.195$, $\eta_p^2=.039$ (three-way interaction). To examine the evidence in support of the null hypothesis for the abovementioned three-way interaction (load x secondary task x age), a Bayesian ANOVA was conducted using JASP (JASP Team, 2018). A Bayesian ANOVA with default priors suggested that the data were 4.193 times more likely to occur under the model that did not include the three-way interaction, compared to the model that did include the three-way interaction, providing positive evidence (Raftery, 1995) for the null-hypothesis (that the interaction between secondary task type and working memory load is not modulated by age).

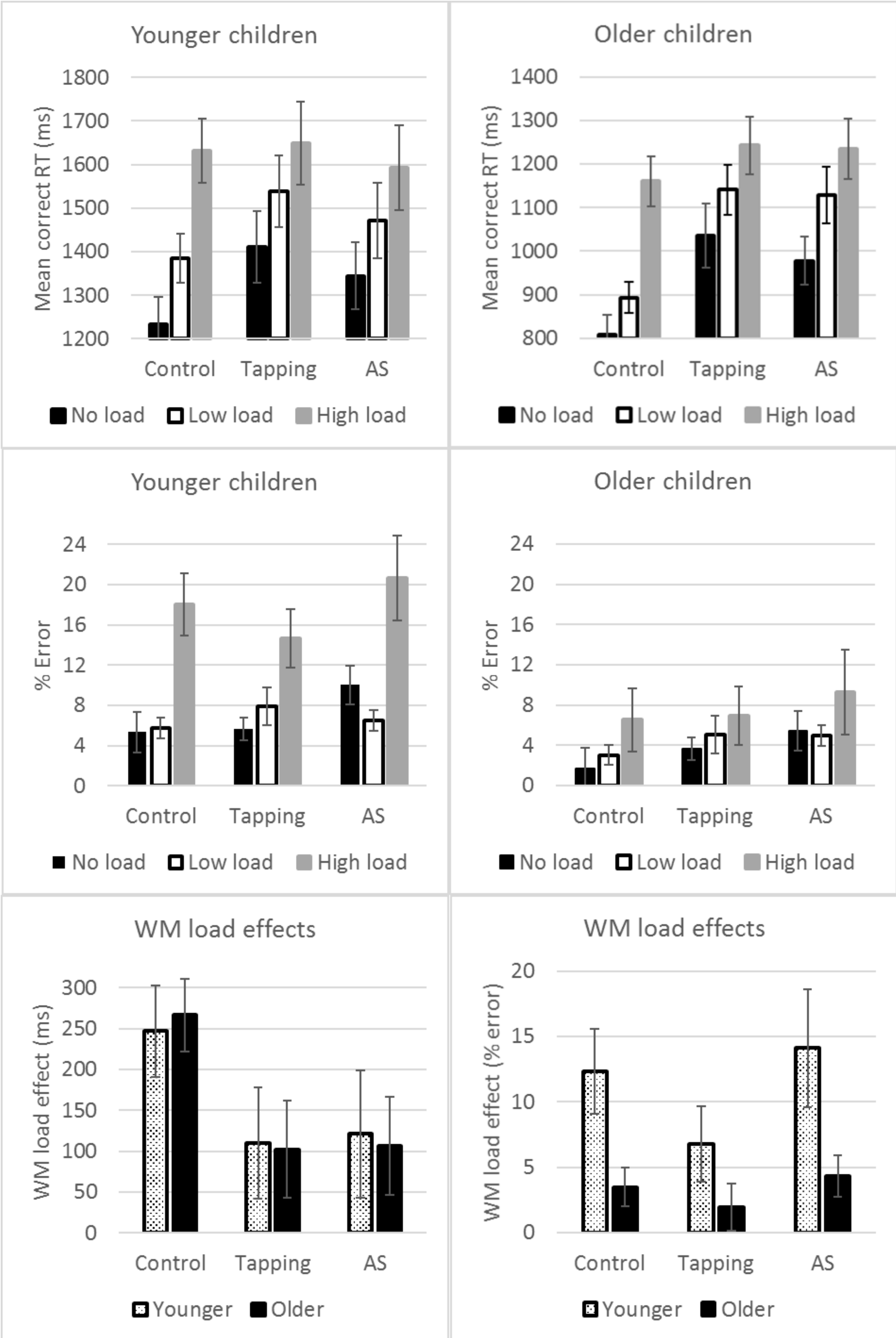
Younger children
Mean correct RT (ms)
1800
1700
1600
1500
1400
1300
1200
Control
Tapping
AS
No load
Low load
High load
Older children
Mean correct RT (ms)
1400
1300
1200
1100
1000
900
800
Control
Tapping
AS
No load
Low load
High load
Younger children
% Error
24
20
16
12
8
4
0
Control
Tapping
AS
No load
Low load
High load
Older children
% Error
24
20
16
12
8
4
0
Control
Tapping
AS
No load
Low load
High load
WM load effects
WM load effect (ms)
300
250
200
150
100
50
0
Control
Tapping
AS
Younger
Older
WM load effects
WM load effect (% error)
20
15
10
5
0
Control
Tapping
AS
Younger
Older

Figure 8 Top/middle panel: Mean correct RTs ± SEM (top) and % error ± SEM data (middle) for younger children (left) and older children (right) in Experiment 3, as a function of working memory load (none, low or high) and secondary task type (control, foot tapping or articulatory suppression). Bottom panel: Working memory load effects for younger and older children as a function of secondary task type (control, foot tapping or articulatory suppression) for the mean correct RT (left) and % error (right) data.

Summary

This experiment investigated what causes the age difference in working memory capacity for task rules. Younger (Year 1) and older (Year 4 and 5) children completed a stimulus classification task in which the working memory load was either high (eight S-R rules) or low (two S-R rules), whilst simultaneously performing a secondary task (AS or FT), or not. If the age difference in working memory capacity (as measured by the working memory load effect) is strategic; then this age difference should be reduced or eliminated when the use of maintenance mechanisms is prevented by the addition of a secondary task. However, the data argue against this possibility.

In the reaction time data, the working memory load effect was reduced by the addition of a secondary task, but equally so for younger and older children: For both age groups, performance in the low load condition was affected more by a secondary task than performance in the high load condition. One possible explanation for this result is that in the low load condition, the two relevant task rules are within working memory capacity; and therefore any other (secondary) task will impact on performance; whereas in the high load condition, because not all eight task rules could be maintained within working memory on a given trial (hence requiring long term memory retrieval on some trials, even without a secondary task), the effect of a concurrent task is less severe.

One surprising result from Experiment 3 was that, unlike in Experiment 1, the RT working memory load effect did not significantly differ as a function of age[4]. However, a significant interaction between age and working memory load was present in the error data. This (manifestation of an effect in the error but not the RT data) is not unusual, as accuracy has previously been argued to be a more sensitive measure than RT in young children (Diamond & Kirkham, 2005). Additionally, the extent to which participants prioritize speed or accuracy is known to be under strategic control (Rinkenauer, Osman, Ulrich, Müller-Gethmann & Mattes, 2004), and so it is possible that the age difference in the working memory load effect manifested itself in the RT data in Experiment 1 and in the accuracy data in Experiment 3 because of strategic reasons. The most important finding, though, is that an age difference in the working memory load effect was present in both Experiments 1 and 3.

In the error data, the working memory load effect also appeared larger under AS than under tapping, though this result is difficult to interpret because the effect of working memory load was not larger under tapping compared to the control condition; and because AS was the more difficult secondary task even in the absence of a memory load. However, the key result is that the error data analysis found no evidence that the age difference in working memory load effect could be reduced or eliminated by the presence of a secondary task (indeed, a Bayesian ANOVA provided positive evidence in support of the null-hypothesis) This demonstrates that the age difference in working memory capacity is not the result of older children's ability to use the same resources more efficiently. Rather, these results suggest that a structural difference (i.e. older children have increased working memory capacity) is the cause of the observed age difference in the effect of working memory load.

Discussion

[4] Note that for this reason (lack of interaction between age and working memory load in the mean correct RT data), further proportional analyses were not conducted on the RT data of Experiment 3.

Definitions of working memory clearly emphasize its importance in regulating goal-driven behavior. Yet the majority of working memory research focusses on our ability to recall discrete items of information; and much less is known about how working memory enables us to act on this information. This study investigated children's working memory capacity for procedural information (specifically task rules), and how it develops throughout childhood. Experiment 1 varied the number of stimulus-response (S-R) rules in a choice RT paradigm (using the set-size effect described by Hick's law as an index of working memory capacity) and found that the RT set-size effect decreased with age. A proportional RT analysis, together with Experiment 2, confirmed that this observed developmental difference in working memory capacity is unlikely to be caused by overall age differences in RT. Finally, Experiment 3 replicated the age difference in working memory capacity for task rules using novel tasks and stimuli, whilst also demonstrating that this age difference in working memory performance cannot be explained in terms of age differences in the use of (verbal) strategies. Instead, a basic difference in available capacity to actively maintain S-R mappings appears to be at the root of this developmental difference.

The experiments reported here constitute the first convincing evidence for a developmental increase in working memory capacity for task rules. Below we will firstly discuss how these results relate to previous findings. We will then discuss in more detail both the causes of this capacity difference, and the likely consequences of this developmental difference for other cognitive processes (such as cognitive skill acquisition, and task switching performance) which depend in part on working memory capacity.

The age-related increase in working memory capacity for task rules is consistent with some previous findings, but not others. Specifically, Kiselev et al. (2009) also found the set-size effect to decrease with age. Our findings support these results, and furthermore suggest that Davidson et al.'s (2006) results (an increase in the RT set-size effect with age) were caused by a speed accuracy trade-off which is apparent in their data. Moreover, in contrast to these previous experiments, the current study ensured that the frequency of stimuli was matched between the set-size conditions. This is vitally important, as uncontrolled data sets are known to result in inflated set-size effects (Van 't Wout, 2018). Specifically, it is possible that the influence of stimulus frequency on the set-size effect differs as a function of age. For example, if younger children's performance is more affected by stimulus frequency, then the previously found inverse relationship between set-size effect and age (Kiselev et al., 2009) could have been caused by a stimulus frequency confound, rather than an age difference in working memory capacity.

It is worth noting that by varying the number of stimulus-response mappings, Experiments 1 and 2 not only manipulated working memory load, they also manipulated the number of response options (spaceships) that the participant had to attend to, and select from. One might therefore ask whether the larger set-size effect observed for the younger children instead reflects a reduced ability to attend to one response option amongst alternatives, or simply a reduced ability to execute one motor response amongst alternatives. The results of Anderson, Nettelbeck and Barlow (1997), who used a 'Jensen procedure' to investigate the development of speed of processing, argue against the latter possibility. Specifically, in the Jensen procedure, participants are required to press the response button position immediately below one of several stimulus lights. These stimulus lights are arranged in a semi-circle, equidistant from a central "home button". On each trial, participants hold down the home button, and then press the response button immediately underneath the illuminated stimulus light. Using this set-up, Anderson et al. (1997) tested 7- and 11-year-old children in three conditions: with 2, 4 and 8 stimulus lights and found no reliable interaction between number of stimulus lights and age. This finding suggests that the observed interaction between age and set-size in the current study does reflect the differential effects of memory load on the different age groups, rather than the effects of any increases in attentional and motor demands with set size. Also consistent with this view are the results of Experiment 3, in which an age difference in working memory capacity was found even though, across participants, the same stimuli and response options were used in both working memory load conditions.

The next obvious question, which was addressed by Experiment 3, concerns the potential causes of the observed difference in working memory capacity. Is the observed age difference in working memory performance caused by an age difference in strategic ability (resulting from either a production or a utilization deficiency; Miller 1994), or by an increase in available capacity? Previous studies addressing this question have yielded mixed results (e.g., Tam et al., 2010; Cowan et al., 2011). Additionally, all of those studies assessed children's ability to maintain declarative representations in working memory; and it remains unknown whether the same strategies can also be used to maintain procedural representations in working memory.

To examine whether age difference in strategic (specifically: verbal) ability are at the root of the observed difference in working memory capacity, Experiment 3 compared the working memory load effect in two different age groups, under three secondary task conditions: a verbal distractor task (articulatory suppression), a nonverbal distractor task (foot tapping) or no secondary task. If the increase with age in working memory capacity for task rules (demonstrated by Experiments 1 and 2) was caused by older children's more efficient use of a (verbal or nonverbal) maintenance mechanism, then the age difference in working memory capacity should be modulated by the addition of a secondary task (cf. Cowan, Nugent, Elliot, Ponomarev & Saults, 1999). However, the results argued against this possibility: although the working memory load effect was affected by the addition of a secondary task, this interaction was not further modulated by age. Indeed, with regards to the three-way interaction between age, secondary task and working memory load, a Bayesian ANOVAS provided positive evidence in support of the null-hypothesis.

The results of Experiment 3 are therefore consistent with a developmental difference in basic working memory capacity for task rules. This "capacity account" can be further subdivided into i) theories which assume that working memory has a fixed number of slots (Cowan, 2001; Miller, 1956) and ii) theories which assume that the limited capacity of working memory is the result of a limited resource of distributed activity (Ma, Husain & Bays, 2014). According to the former view, age-related improvements in working memory are caused by an increase in the number of available slots. Although the latter view has not been explicitly applied to the developmental literature, it would predict that age-related improvements in working memory performance can be explained in terms of an increase in the total "pool" of activation available to be distributed.

These two characterizations of working memory capacity (slots vs. spreading activation) make distinct predictions about the data reported here. Specifically, according to the activation-based account, RT should steadily increase with the number of S-R mappings. Conversely, according to the slot-based account, RT would increase non-linearly, peaking when the number of slots has been filled, and plateauing thereafter. Hence, the slot-based account would predict an age difference in the *shape* of the function, rather than the *slope* of the function. At first sight, in this study RT appeared to increase linearly with the number of S-R mappings for all age groups. Hence, the mean RT data appear to be more consistent with an activation-based account of working memory capacity. However, it is possible that the plateau predicted by the slot-based account would be observed if children were tested using a greater number of S-R mappings. Furthermore, the RT distribution analysis, which showed similar set-size effects for the age groups at the fast end of the RT distribution, was more consistent with the latter (slot-based) account: The increase in set-size effect with age for slow responses suggests that, with a reduced number of "slots", younger children were more likely *not* to have to relevant S-R rule in working memory when the number of potentially relevant S-R rules exceeded capacity. In contrast, the lack of an effect of age on set-size for fast responses suggests that, for a proportion of trials in the high load (5 S-R) condition, even the youngest children had the correct S-R rule in working memory. The results of this distribution analysis were therefore more consistent with a slot-based account of working memory capacity, and argue against an activation-based account, according to which the set-size effect should be comparable throughout the distribution.

Finally, one important question which remains unanswered is whether the ability to maintain and implement procedural representations is distinct from the ability to maintain declarative representations. As mentioned in the introduction, most theories of working memory do not incorporate this distinction, and the two studies which have addressed this issue (Barrouilet et al., 2015; Gade et al., 2014) have yielded inconclusive results. A direct comparison between declarative and procedural representations is complex, for at least two reasons: First, although it is relatively straightforward to assess a person's ability to maintain declarative representations without the involvement of an obvious procedural component (using span tasks, for example), the reverse is much more difficult, as most procedural representation can still be described in declarative terms. Secondly, as noted by Gade et al. (2014), the paradigms used to study working memory for procedural and declarative representations are very different. For example, simple and complex span experiments, which are used to study the maintenance of declarative information, usually present subjects with a new list of items on each trial. Conversely, task switching experiments, which capture the storage and manipulation of procedural information, typically involve the same procedural representations for hundreds of trials throughout the experiment. Attempts to compare procedural and declarative working memory are complicated by both of these factors.

Despite this, studies using the goal neglect procedure (Duncan et al., 1996) and the DCCS (Zelazo et al., 1996), as described in the Introduction, have demonstrated that we must on some level distinguish between a person's ability to remember a (declarative) representation, and a person's ability to implement the associated (procedural) representation. The phenomenon of goal neglect demonstrates just this: It is possible for people to correctly remember a rule, yet be unable to implement this rule. Although this clearly demonstrates a distinction between recalling a piece of information and acting on that information, the cognitive mechanism underlying this distinction remains unclear. Specifically, it remains unknown whether separate subsystems of working memory are responsible for the storage of procedural and declarative information, or whether a single working memory supports the maintenance of both declarative and procedural representations. Future research will need to distinguish between these accounts. Regardless of the outcome, given the findings of Duncan et al. (1996) and Zelazo et al. (1996) described above, it is vital that working memory research investigates not only people's ability to maintain information for a short period of time, but also their ability to act on that information.

Summary and future directions

To summarize, the experiments reported here sought to determine whether children's ability to actively maintain procedural representations in working memory improves throughout childhood. The three experiments reported here have shown that working memory capacity for procedural representations increases with age. A proportional analysis, combined with the results of Experiment 2 (in which an older group of children performed a more difficult version of the task) demonstrated that this result is unlikely to be driven by overall RT differences between the age groups. Finally, Experiment 3 indicated that the age difference in working memory capacity for task rules is not the result of strategic differences between the age groups, but rather reflects a basic difference in capacity. Together, these results imply that the age difference in working memory capacity for task rules is likely to have important consequences for behavior. As working memory is thought to make an important contribution to other crucial cognitive processes, such as the acquisition of novel skills, and executive control, it is highly likely that this developmental difference in the ability to represent and implement task rules further impacts on such cognitive processes. Indeed, in a companion paper we have already demonstrated that children's working memory capacity for task rules crucially affects task switching performance (Van 't Wout, O'Donnell, Saw & Jarrold, in preparation). These results clearly demonstrate that the developmental increase in working memory capacity has important consequences beyond the recall of lists of items. Further research must focus on these consequences in order to gain a comprehensive account of the many ways in which working memory capacity affects not only children's recall ability, but also their behavior.

References

Anderson, M., Nettelbeck, T., & Barlow, J. (1997). Reaction time measures of speed of processing: Speed of response selection increases with age but speed of stimulus categorization does not. *British Journal of Developmental Psychology, 15*, 145-157.

Barrouillet, P., Corbin, L., Dagry, I., & Camos, V. (2015). An empirical test of the independence between declarative and procedural working memory in Oberauer's (2009) theory. *Psychonomic Bulletin & Review, 22 (4)*, 1035-1040.

Barrouillet, P, Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology, 45*, 477-490.

Bayliss, D., Jarrold, C., Baddeley, A., Gunn, D., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology, 41*, 579-597.

Brass, M., Liefooghe, B., Braem, S., & De Houwer, J. (2017). Following new task instructions: Evidence for a dissociation between knowing and doing. *Neuroscience and Biobehavioral Reviews, 81*, 16-28.

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language, 61*, 457-469.

Case, R., Kurland, D.M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-114.

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. (2011). Age differences in visual working memory capacity: Not based on encoding limitations. *Developmental Science, 14 (5)*, 1066-1074.

Cowan, N., Cartwright, C., Winterowd, C., & Sherk, M. (1987). An adult model of preschool children's speech memory. *Memory & Cognition, 15*, 511–517.

Cowan, N., Nugent, L.D., Elliott, E.M., Ponomarev, I., & Saults, J.S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70*, 1082-1097.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Davidson, M. C., Amso, D., Anderson, L. C., Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44 (11)*, 2037-2078.

Diamond, A., & Kirkham, N (2005). Not quite as grown-up as we like to think: Parallels between cognition in childhood and adulthood. *Psychological Science, 16(4)*, 291-297.

Duncan, J., Emslie, H., Williams, P., & Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: the organization of goal directed behavior. *Cognitive Psychology, 30*, 257-303.

Duncan, J., Schramm, M., Thompson, R., & Dumontheil, I. (2012). Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin & Review, 19*, 864-870.

Gade, M., Druey, M. D., Souza, A. S., & Oberauer, K. (2014). Interference within and between declarative and procedural representations in working memory. *Journal of Memory and Language, 76*, 174-194.

Gaillard, V., Barrouillet, P., Jarrold, C., & Camos, V. (2011). Developmental differences in working memory: Where do they come from? *Journal of Experimental Child Psychology, 110*, 469-479.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Science, 3*, 410-418.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11-26.

Hommel, B. (1998). Automatic stimulus-response translation in dual task performance. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1368-1384.

Hübner, M., Kluwe, R. H., Luna-Rodriguez, A., & Peters, A. (2004). Response selection difficulty and asymmetrical costs of switching between tasks and stimuli: No evidence for an exogenous component of task-set reconfiguration. *Journal of Experimental Psychology: Human Perception and Performance, 30* (6), 1043-1063.

Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory and Cognition, 35*, 1675-1684.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38*, 241-253.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 53*, 188-196.

JASP Team (2018). JASP (Version 0.9)[Computer software].

Kiselev, S., Espy, K. A., & Sheffield, T. (2009). Age related differences in reaction time performance in young children. *Journal of Experimental Child Psychology, 102*, 150-166.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30.000 English words. *Behavioral Research Methods, 44 (4),* 978-990.

Kveraga, K., Boucher, L., & Hughes, H. C. (2002). Saccades operate in violation of Hick's law. *Experimental Brain Research, 146*, 307-314.

Ma, W. J., Husain, M., Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17 (3)*, 347-356.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Muhle-Karbe, P. S., Duncan, J., De Baene, W., Mitchell, D. J., & Brass, M. (2016). Neural coding for instruction-based task sets in human frontoparietal and visual cortex. *Cerebral Cortex, 27 (3)*, 1891-1905.

Oberauer, K. (2009). Design for a working memory. *Journal of Learning and Motivation, 51*, 45-100.

Oberauer, K. (2010). Declarative and procedural working memory: common principles, common capacities? *Psychologica Belgica, 50 (3-4)*, 277-308.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. M. (2005). Working memory and intelligence – their correlation and their relation: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 61-65.

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods, 162 (1-2)*, 8-13.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), Sociological methodology 1995 (pp. 111–196). Cambridge, MA: Blackwell.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114 (3), 510-532.

Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the locus of speed–accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General, 133 (2)*, 261-282.

Schneider, D. W., & Anderson, J. R. (2011). A memory-based model of Hick's law. *Cognitive Psychology, 62*, 193-222.

Tam, H., Jarrold, C., Sabatos-DeVito, M., & Baddeley, A.D. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. *Journal of Experimental Child Psychology, 107*, 306-324.

Towse, J. N., Lewis, C., & Knowles, M. (2007). When knowledge is not enough: The phenomenon of goal neglect in preschool children. *Journal of Experimental Child Psychology, 96 (4)*, 320-332.

Van 't Wout, F. (2018). The contribution of stimulus recency and frequency to set-size effects. *Psychonomic Bulletin & Review, 25 (3),* 1123-1128

Van 't Wout, F., Lavric, A., & Monsell, S. (2015). Is it harder to switch among a larger set of tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41 (2)*, 363-376.

Van 't Wout, F., Lavric, A., & Monsell, S. (2013). Are stimulus-response rules represented phonologically for task-set preparation and maintenance*? Journal of Experimental Psychology: Learning, Memory and Cognition, 39*, 1538-1551.

Van 't Wout, F., O'Donnell, M., Saw, R., & Jarrold, C. (in preparation). Is the effect of age on task switching performance modulated by working memory load? An investigation into the relationship between working memory capacity and executive control.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37-63.